



Estimating Cointegrating Relations from a Cross Section

Madsen, Edith

Publication date:
2004

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Madsen, E. (2004). *Estimating Cointegrating Relations from a Cross Section*. Department of Economics, University of Copenhagen.



CAM

**Centre for Applied
Microeconometrics**

**Institute of Economics
University of Copenhagen**

<http://www.econ.ku.dk/CAM/>

Estimating Cointegrating Relations from a Cross Section

Edith Madsen

2004-21

Estimating cointegrating relations from a cross section

Edith Madsen[†]

Institute of Economics, University of Copenhagen, Denmark

November 2004

Abstract

This paper specifies a regression model describing cointegrating relations between variables at the individual level. The models considered allow for homogeneous cointegration and heterogeneous cointegration. In both cases correlation between the regressors and the regression error can occur through aggregate shocks that are common to all cross-section units so the condition about the regressors being independent of the regression error is not imposed. It is shown that the estimator obtained by a cross-section regression performed at any point in time is a consistent estimator of the cointegrating parameters in the homogeneous case and of the cointegrating parameter means in the heterogeneous case. In both cases the limiting distribution of the cross-section estimator is normal.

Keywords: Dynamic panel data models; Non-stationary panel data; Cointegrating relations; Cross-section regression

JEL classification: C31; C32

[†] Institute of Economics, University of Copenhagen, Studiestræde 6, 1455 Kbh. K., Denmark
Email: edith.madsen@econ.ku.dk

1 Introduction

In this paper we specify a dynamic panel data model for the behavior over time of variables at the individual level. The model expresses that the variables are non-stationary and cointegrated when viewed as time series. Under certain restrictions a cross-section regression, based on observations of cross-section units at one point in time, provides a consistent estimator of the long-run cointegrating relations. So the paper can be seen as establishing a link between the idea that cross-section data contains information about the long-run relationships between variables and the finding that certain relations between economic variables at the aggregate level are stable in the long run even though the variables themselves exhibit non-stationary behavior over time. An important result in the cointegration analysis of variables at the aggregate level is that many variables are endogenous. Therefore it seems natural to allow for endogeneity in the model for the variables at the individual level through aggregate shocks that are common to all cross-section units.

The issue considered in the present paper is related to the fast growing literature on unit roots and cointegration in panel data which started in the beginning of the 1990's. Most of the research within this area has been concerned with asymptotic properties in non-stationary panel data models where both the cross-section dimension (N) and the time-series dimension (T) are large. For reviews of many of the contributions within this research area see Banerjee (1999) and Baltagi & Kao (2000). It is different from most of the previous research performed in panel data models as this has concentrated on asymptotic properties in stationary panel data models where the cross-section dimension is large and the time-series dimension is small. Surveys of most of the research performed within this framework can be found in Hsiao (1986), Mátyas & Sevestre (1992), Baltagi (2001) and Arellano (2003) of which the latter two also consider the analysis of non-stationary panel data.

Many economic relationships are dynamic in nature and within the framework of a dynamic panel data model the best situation is clearly to have observations of many cross-section units over a long period of time. Although more and more such panel data sets become available there are still cases where pure cross-section data (observations of cross-section units at one point in time) provides the main source of information. An example is the analysis of the relation between consumption and income based on household expenditure surveys. In most countries - the exceptions being UK and US - these surveys only cover a short period of time thereby ruling out the possibility of constructing pseudo-panels by following cohorts over a longer period of time. The question is what a cross section obtained at some point in time reflects within a dynamic framework. For instance, do cross-section estimates reflect long-run or short-run relations? Obviously, one cross section does not provide any information on the specific dynamic properties, therefore an explicit dynamic model is needed in order to answer such questions. The first contribution to this discussion is the paper by Grunfeld (1961). The paper considers a specific dynamic panel data model (a partial adjustment model) and investigates what a cross-section regression will reveal. In particular it gives conditions for a cross-section regression to provide a consistent estimator of the long-run relations. The paper was written long before long-run relations were given a natural

interpretation in terms of cointegration within a non-stationary framework. Recently the papers by Adda & Robin (1998, 2003) have taken up the discussion again. They investigate estimators obtained by cross-section regression in the case where the variables are described by unit root processes and apply the idea in the estimation of demand systems based on household expenditure survey data from France.

The starting point in the present paper is a regression model describing homogeneous cointegrating relations between variables at the individual level. Suppose that the $I(1)$ regressor is generated independently of the stationary regression error and that innovations are iid across cross-section units. In this case a cross-section regression performed at any point in time provides a consistent estimator (as $N \rightarrow \infty$) of the parameters in the cointegrating relations. However in the time-series analysis of cointegrated variables at the aggregate level the assumption about the regressor being independent of the regression error is usually regarded as being too strong. It is clear that in the case with unrestricted correlation between the $I(1)$ regressor and the stationary regression error a cross-section regression leads to an inconsistent estimator (as $N \rightarrow \infty$) of the cointegrating parameters. Adda & Robin (1998, 2003) show that the asymptotic bias tends to zero as the point in time where the cross section is obtained tends to infinity. This results from the variance of the $I(1)$ regressor being of higher order than the covariance between the regressor and the regression error when viewed as a function of the point in time and this is well-known from the time-series analysis. The interpretation of the result is that the asymptotic bias will be small when the processes describing the time-series behavior of the variables have started a long time before the cross-section data is obtained. However, the result does not provide any information about the actual size of the asymptotic bias, or to put it differently the asymptotic bias depends on time-series properties which can not be identified from one observation of cross-section units. Therefore the result in Adda & Robin (1998, 2003) is not very useful in practice when we are interested in making inference on the cointegrating parameters.

In similar models for panel data with large N and large T , the assumption about the regressors being independent of the regression error is not needed in order to obtain a consistent estimator of the cointegrating parameters. Phillips & Moon (1999) show that a modified pooled estimator obtained by using techniques from the time-series analysis to correct for correlation between the regressor and the regression error is consistent (as $N \rightarrow \infty$ and $T \rightarrow \infty$). With only one observation of the cross section it is clearly not possible to use such techniques. The present paper shows that the assumption about the regressors being independent of the regression error can be weakened in order to allow for some degree of correlation between the regressor and the regression error. It is done by allowing for correlation to occur through aggregate shocks that are common to all cross-section units and therefore can be removed by subtracting the cross-section sample mean from all variables. When this type of endogeneity is allowed in the regression model described above, the paper shows that the usual asymptotic results (as $N \rightarrow \infty$) well-known from ordinary regression theory apply to the estimator obtained by a cross-section regression. In particular the estimator is a consistent estimator of the cointegrating parameters and its limiting distribution is normal. This means that the $I(1)$ property of the regressor will not affect the

inference procedure itself but the asymptotic variance of the cross-section estimator will depend on the point in time where the cross-section data is obtained. Similar results are shown for the model with randomly different cointegrating parameters. In this case a cross-section regression provides a consistent estimator of the cointegrating parameter means. This result is shown in the paper by Pesaran & Smith (1995) under the very strong assumption that the regressors are independent of the regression error. With respect to the $I(1)$ property of the regressor, the effect on the asymptotic variance of the cross-section estimator will be different from that in the case with homogeneous cointegration.

The paper is organized in the following way. Section 2 introduces the basic model and gives the underlying assumptions. Section 3 derives the asymptotic properties of the estimator obtained by a cross-section regression. Section 4 derives the asymptotic properties of the cross-section estimator in an extended version of the basic model with randomly different cointegrating parameters. In Section 5 the main results are illustrated in a simulation study. Finally, Section 6 concludes the paper.

2 The basic model and assumptions

We consider the variables $Y_{it}(k_0 \times 1)$, $X_{1it}(k_1 \times 1)$ and $X_{2it}(k_2 \times 1)$ where $i = 1, \dots, N$ and $t = 1, 2, \dots$. For every cross-section unit $i = 1, \dots, N$ we assume that the variables are generated by the following model

$$Y_{it} = \gamma_1' X_{1it} + \gamma_2' X_{2it} + \eta_{0it} \quad (1)$$

$$X_{1it} = \eta_{1it} \quad (2)$$

$$X_{2it} = X_{2it-1} + \eta_{2it} \quad (3)$$

where γ_1 and γ_2 are $k_1 \times k_0$ and $k_2 \times k_0$ matrices of parameters, respectively, and where the time-series processes η_{0it} , η_{1it} and η_{2it} are (weakly) stationary for every cross-section unit $i = 1, \dots, N$.

Under additional assumptions concerning the time-series properties of η_{0it} , η_{1it} and η_{2it} that we will give below, the model expresses the following. Viewing the variables Y_{it} , X_{1it} and X_{2it} as time series, the variables Y_{it} and X_{2it} are $I(1)$ whereas the variable X_{1it} is stationary. Furthermore the two $I(1)$ variables Y_{it} and X_{2it} are cointegrated where the cointegrating relations are described by equation (1). Equation (2) can also be considered as a cointegrating relation expressing that the variable X_{1it} is trivially cointegrated since this variable is stationary. Altogether this means that γ_2 describes a long-run or equilibrium relation between the two non-stationary variables Y_{it} and X_{2it} which is the same for all cross-section units. Notice that γ_1 does not have a similar interpretation. In fact the parameter γ_1 is not identified in the time-series sense as $Y_{it} - \gamma_2' X_{2it}$ plus any linear combination of the variable X_{1it} is stationary.

When viewing the variables Y_{it} , X_{1it} and X_{2it} as time series, the model is of the same type as in Park & Phillips (1989). In line with their work, equation (1) can be considered as a regression equation where the assumption about the regressors X_{1it} and X_{2it} and the regression error η_{0it} being independent of each other is not imposed. Within this framework Park & Phillips (1989) show that it is possible to

obtain a (super)consistent estimator of the long-run relation, i.e. the parameter matrix γ_2 , from a time-series regression with or without including X_{1it} as a regressor. However in order to obtain a consistent estimator of the long-run relation from a cross-section regression the correlation between the regressors and the regression error must be restricted. In particular, this means that it might be necessary to include the stationary variable X_{1it} as a regressor. We will return to this issue in the following.

In order to specify the behavior of the variables further we assume that the terms η_{0it} , η_{1it} and η_{2it} can be decomposed in the following way

$$\eta_{0it} = \mu_{0i} + u_{0t} + v_{0it} \quad (4)$$

$$\eta_{1it} = \mu_{1i} + u_{1t} + v_{1it} \quad (5)$$

$$\eta_{2it} = \mu_{2i} + u_{2t} + v_{2it} \quad (6)$$

μ_{0i} , μ_{1i} and μ_{2i} describe individual-specific effects that are constant over time. u_{0t} , u_{1t} and u_{2t} represent time-specific effects that are the same for all cross-section units. Finally v_{0it} , v_{1it} and v_{2it} describe individual-specific effects that are time-dependent. To be more explicit about the nature of the terms in the expressions (4)-(6) the assumptions listed below are imposed.

Assumption 1 *The terms $\mu_i = (\mu'_{0i}, \mu'_{1i}, \mu'_{2i})'$ where $i = 1, \dots, N$ are independent and identically distributed across i with finite second order moment.*

Note that μ_i is not assumed to have mean zero. The term μ_{1i} allows for individual-specific effects in the level of the variable X_{1it} that are constant over time. The term μ_{2i} has a similar effect on changes in the variable X_{2it} meaning that μ_{2i} generates an individual-specific linear trend in the variable X_{2it} when viewed as a time series. Finally the term μ_{0i} allows for individual-specific deviations from the long-run relation given by (1) that are constant over time.

Assumption 2 *The terms $v_{it} = (v'_{0it}, v'_{1it}, v'_{2it})'$ where $i = 1, \dots, N$ are independent and identically distributed across i with mean zero and finite second order moment for all $t = 1, 2, \dots$. Furthermore, for every $i = 1, \dots, N$ the time-series process v_{it} is weakly stationary.*

The assumption allows for short-run heterogeneity of the type considered in Phillips & Moon (1999) where v_{it} is a linear process with random coefficients, i.e. $v_{it} = \sum_{s=0}^{\infty} C_{is} \varepsilon_{it-s}$ where C_{is} is iid across i and ε_{it} is iid across i, t . So the dynamic structure of v_{it} is to some extent allowed to vary across units even though v_{it} is iid across i . Phillips & Moon (1999) show that if the v_{it} 's are generated by random coefficient ARMA processes where the roots of the characteristic equations are in a compact set inside the unit circle, then the second order moments of v_{it} are well-defined. When v_{2it} in addition to being stationary is $I(0)$ when viewed as a time series there are individual-specific $I(1)$ trends in the variable X_{2it} and by that in the variable Y_{it} when these are viewed as time series.

Assumption 3 *For every cross-section unit $i = 1, \dots, N$ the following hold:*

- (i) μ_{0i} is independent of μ_{1i} and μ_{2i}
- (ii) v_{0it} is generated independently of v_{1it} and v_{2it}

The assumption above implies that no correlation between the regressors X_{1it} and X_{2it} and the regression error η_{0it} is allowed through the individual-specific terms. In the case where $v_{it} = \sum_{s=0}^{\infty} C_{is}\varepsilon_{it-s}$ then (ii) means that the first k_0 elements and the last $(k_1 + k_2)$ elements in ε_{it} are independent of each other and that the matrices C_{is} are all block diagonal where the first block is of dimension k_0 and the second block is of dimension $(k_1 + k_2)$. However the assumption about the regressors being uncorrelated with the regression error is usually regarded as being too strong in the literature on time-series analysis of cointegrated variables at the aggregate level. Therefore it seems natural to allow for correlation between the regressors and the regression error through aggregate terms that are common for all cross-section units.

Assumption 4 *The process $u_t = (u'_{0t}, u'_{1t}, u'_{2t})'$ is weakly stationary with mean zero. In addition $\sum_{s=-\infty}^{s=\infty} \Gamma_{2u}(s)$ is positive definite where $\Gamma_{2u}(s)$ is the autocovariance function of the process u_{2t} .*

The term u_t represents aggregate shocks that are common to all units. As mentioned above we allow for some degree of endogeneity between the regressors X_{1it} and X_{2it} and the dependent variable Y_{it} through this term since the components in u_t are not assumed to be uncorrelated. It is very important for the results in the next section that it is possible to remove the aggregate shocks and by that the endogeneity between the regressors and the dependent variable by subtracting the cross-section sample mean from all variables. This means that no individual-specific reaction to the aggregate shocks is allowed in the model. All cross-section units must react in precisely the same manner to aggregate shocks and this is of course a quite strong assumption. In addition the term u_t introduce the most simple form of dependency between the cross-section units. The assumption on the autocovariance function of u_{2t} ensures that all components of X_{2it} are $I(1)$ and that there are no cointegrating relations between these components, see Phillips (1986). This implies that the model defined by the equations (1)-(3) is indeed what it seems to be. In particular the parameter matrix γ_2 describes a long-run relation between the two non-stationary variables Y_{it} and X_{2it} .

Also an assumption concerning the initialization of the variable X_{2it} is needed. The assumption is the following.

Assumption 5 *X_{2i0} where $i = 1, \dots, N$ are independent and identically distributed across i with finite second order moment.*

Finally, we need an assumption about the joint behavior of the different terms generating the variables Y_{it} , X_{1it} and X_{2it} .

Assumption 6 *μ_i , v_{is} , u_t and X_{2i0} are mutually independent for all $i = 1, \dots, N$ and all $s, t = 1, 2, \dots$.*

Altogether we have a linear regression model describing cointegrating relations between the two non-stationary variables Y_{it} and X_{2it} . The terms in X_{2it} which cause the non-stationarity in the time-series sense are first of all $I(1)$ trends that are common to all cross-section units and an individual-specific linear trend. In addition to these terms there might be individual-specific $I(1)$ trends as well. An important feature of the model is that correlation between the regressors X_{1it} and X_{2it} and the regression error η_{0it} is allowed through terms that are common to all cross-section units. This means that all stationary variables X_{1it} where correlation with the $I(1)$ regressor X_{2it} occur through individual-specific terms must be included as regressors in the model.

3 Cross-section regression

We consider the dynamic model specified in the previous section as a model for the variables at the individual level. Suppose a cross section obtained at some point in time is available. The cross section is a sample consisting of observations of N cross-section units at time t where $t \in \mathbb{N}$. First of all the cross-section sample mean is subtracted from all variables and the corrected variables are defined as $Y_{it}^* = Y_{it} - \frac{1}{N} \sum_{i=1}^N Y_{it}$, $X_{1it}^* = X_{1it} - \frac{1}{N} \sum_{i=1}^N X_{1it}$ and $X_{2it}^* = X_{2it} - \frac{1}{N} \sum_{i=1}^N X_{2it}$. For notational convenience the stacked $(k_1 + k_2)$ -dimensional stochastic variable X_{it}^* is defined as $X_{it}^* = (X_{1it}^{*'}, X_{2it}^{*'})'$ and the corresponding $(k_1 + k_2) \times k_0$ parameter matrix as $\gamma = (\gamma_1', \gamma_2')'$. Now the regression equation in (1) describing the cointegrating relations can be expressed in terms of the corrected variables in the following way

$$Y_{it}^* = \gamma' X_{it}^* + \eta_{0it}^* \quad i = 1, \dots, N \text{ and } t \in \mathbb{N} \quad (7)$$

where $\eta_{0it}^* = \eta_{0it} - \frac{1}{N} \sum_{i=1}^N \eta_{0it}$. The corresponding cross-section ordinary least square estimator denoted $\hat{\gamma}_{N,t}$ is defined as

$$\hat{\gamma}_{N,t} = \left(\sum_{i=1}^N X_{it}^* X_{it}^{*'} \right)^{-1} \left(\sum_{i=1}^N X_{it}^* Y_{it}^{*'} \right) \quad (8)$$

According to Assumption 3 the regressor X_{it}^* is independent of the regression error η_{0it}^* as the aggregate shocks have been removed from the variables. This immediately implies that $\hat{\gamma}_{N,t}$ is an unbiased estimator of γ , i.e. $E(\hat{\gamma}_{N,t}) = \gamma$. The asymptotic behavior as $N \rightarrow \infty$ of the cross-section estimator $\hat{\gamma}_{N,t}$ is given in the proposition below.

Proposition 1 *Under Assumption 1-6 the following hold:*

$\hat{\gamma}_{N,t}$ *is a consistent estimator of* γ , *i.e.*

$$\hat{\gamma}_{N,t} \xrightarrow{P} \gamma \text{ as } N \rightarrow \infty \quad (9)$$

The limiting distribution of $\hat{\gamma}_{N,t}$ *is given by*

$$\sqrt{N} (\hat{\gamma}_{N,t} - \gamma) \xrightarrow{w} N(0, \Omega \otimes \Sigma_t^{-1}) \text{ as } N \rightarrow \infty \quad (10)$$

The asymptotic variance can be estimated consistently by using the following results

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} \xrightarrow{P} \Sigma_t \text{ as } N \rightarrow \infty \quad (11)$$

$$\frac{1}{N} \sum_{i=1}^N (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*) (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \xrightarrow{P} \Omega \text{ as } N \rightarrow \infty \quad (12)$$

The proof of Proposition 1 is given in Appendix A.1. The proposition shows that by using ordinary regression methods it is possible to make asymptotic inference on the cointegrating parameters from a cross section obtained at any point in time. Since the regressor X_{2it} is non-stationary when viewed as a time series, the asymptotic variance of $\sqrt{N}(\hat{\gamma}_{N,t} - \gamma)$ depends on the point in time where the cross section is obtained. Below this property is investigated in detail.

We consider the cross-section estimator of γ_2 defined as the submatrix of $\hat{\gamma}_{N,t}$ corresponding to the regressor X_{2it} . To be more specific let this estimator denoted $\hat{\gamma}_{2,N,t}$ be the last k_2 rows in $\hat{\gamma}_{N,t}$ and let Σ^{22t} be the lower $k_2 \times k_2$ diagonal block matrix of Σ_t^{-1} , i.e.

$$\Sigma^{22t} = (\Sigma_{22,t} - \Sigma_{21,t} \Sigma_{11,t}^{-1} \Sigma_{12,t})^{-1} \quad (13)$$

where Σ_t is decomposed according to X_{1it} and X_{2it} as

$$\Sigma_t = \begin{bmatrix} \Sigma_{11,t} & \Sigma_{12,t} \\ \Sigma_{21,t} & \Sigma_{22,t} \end{bmatrix} \quad (14)$$

Then according to Proposition 1 the limiting distribution of $\hat{\gamma}_{2,N,t}$ is given by

$$\sqrt{N}(\hat{\gamma}_{2,N,t} - \gamma_2) \xrightarrow{w} N(0, \Omega \otimes \Sigma^{22t}) \text{ as } N \rightarrow \infty \quad (15)$$

The following assumption is used in the results given below.

Assumption 7 For $a \in \mathbb{R}$ the diagonal matrix F_t is defined in the following way

$$F_t = \begin{bmatrix} I_{k_1} & 0 \\ 0 & t^a I_{k_2} \end{bmatrix}$$

and the following condition is satisfied

$$\lim_{t \rightarrow \infty} (F_t \Sigma_t F_t) \text{ is positive definite} \quad (16)$$

The assumption implies that when the components of Σ_t are normalized correctly with respect to t then the limit of the normalized matrix $F_t \Sigma_t F_t$ as $t \rightarrow \infty$ is well-defined and positive definite. The results below concern the properties of the asymptotic variance of $\hat{\gamma}_{2,N,t}$ as the point in time where the cross section is obtained goes to infinity. Another way of expressing $t \rightarrow \infty$ is by saying that the time origin of the variables Y_{it} , X_{1it} and X_{2it} when viewed as time series is far away. In our model this means that the behavior of the non-stationary regressor X_{2it} is dominated by either linear trends or $I(1)$ trends. If any of these are individual-specific it means that the cross-section variation of the regressor X_{2it} dominates that of the error term as $t \rightarrow \infty$.

Result 1 Let $\Gamma_v(s)$ be the mean autocovariance function of the weakly stationary time-series process $(v'_{1it}, v'_{2it})'$ specified in Assumption 2 and assume that the mean autocovariances are absolutely summable. Then the following hold:

(a) If Assumption 7 is satisfied with $a = -1$ then

$$\Omega \otimes \Sigma^{22t} = O(t^{-2}) \quad (17)$$

(b) If Assumption 7 is satisfied with $a = -1/2$ then

$$\Omega \otimes \Sigma^{22t} = O(t^{-1}) \quad (18)$$

(c) If Assumption 7 is satisfied with $a = 0$ then

$$\lim_{t \rightarrow \infty} (\Omega \otimes \Sigma^{22t}) \text{ is well-defined} \quad (19)$$

The proof of these results are given in Appendix A.2. The assumption in (a) implies that $\text{Var}(\mu_{2i})$ is positive definite. In the time-series dimension this means that there is an individual-specific linear trend in the variable X_{2it} . From the time-series analysis it is well known that asymptotically as $t \rightarrow \infty$ a linear trend will dominate possible $I(1)$ trends. In the cross-section dimension this means that the variation of X_{2it} is of order t^2 . Combining this with the property that the regression error η_{0it} is stationary implies that the convergence rate of the asymptotic variance is of order t^{-2} . That is for any $\varepsilon > 0$ it is always possible to find a point in time t_0 such that in all subsequent time-periods where $t \geq t_0$ then the asymptotic variance normalized by t^{-2} is within an ε -neighborhood of $\lim_{t \rightarrow \infty} (\Omega \otimes t^2 \Sigma^{22t})$. It means that the asymptotic variance is within a (shrinking) ε -neighborhood of $t^{-2} \lim_{t \rightarrow \infty} (\Omega \otimes t^2 \Sigma^{22t})$ for all $t \geq t_0$. So when t is large enough such that $t^{-2} \lim_{t \rightarrow \infty} (\Omega \otimes t^2 \Sigma^{22t})$ provides a good approximation to the asymptotic variance, the result means that there is a gain in efficiency (asymptotically as $N \rightarrow \infty$) by using a cross section obtained at a later point in time. If we take the point in time t as our reference point and compare it with the latter point in time kt where $k \in \mathbb{N}$, then the relative decrease in the asymptotic variance at time kt compared to at time t is $(1 - 1/k^2)$. The decrease is 75% for $k = 2$ and 93.75% for $k = 4$. Note that this only holds if $t^{-2} \lim_{t \rightarrow \infty} (\Omega \otimes t^2 \Sigma^{22t})$ provides a good approximation to the asymptotic variance and that the result does not give information about how large t should be for this to hold.

The assumption in (b) and (c) implies that $\text{Var}(\mu_{2i}) = 0$ meaning that there is no individual-specific linear trend in the variable X_{2it} when viewed as a time series. A possible linear trend in X_{2it} is the same for all cross-section units and is therefore removed when the cross-section sample mean is subtracted. In addition the assumption in (b) implies that $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s)$ is positive definite where $\Gamma_{2v}(s)$ is the mean autocovariance function of v_{2it} . If the cumulated process $v_{2i1} + \dots + v_{2it}$ generates individual-specific $I(1)$ trends in all components of X_{2it} and these components are not cointegrated, then the individual-specific long-run variance of v_{2it} is positive definite, see Phillips (1986). When the individual-specific long-run variances of v_{2it} are majorized by convergent series (This holds if the v_{2it} 's are generated by random

coefficient ARMA processes with certain restrictions on the roots of the characteristic equations, see above) then $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s)$ is the mean long-run variance of v_{2it} which is positive definite when the individual-specific long-run variances of v_{2it} are positive definite with probability one. The proof in Appendix A.2 shows that $\frac{1}{t} \Sigma_{22,t}$ converges to $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s)$ as $t \rightarrow \infty$. Altogether, the assumption in (b) is satisfied when there are individual-specific $I(1)$ trends in all components of X_{2it} and they do not cointegrate. In this case the cross-section variation of X_{2it} is of order t meaning that the convergence rate of the asymptotic variance is of order t^{-1} which is slower than in (a). As above it means that when comparing the asymptotic variance at time t with that of at time kt where $k \in \mathbb{N}$ then the relative decrease will be $(1 - 1/k)$. The decrease is 50% for $k = 2$ and 75% for $k = 4$. So the relative gain in efficiency (asymptotically as $N \rightarrow \infty$) by using a cross section obtained at a later point in time is lower than in (a). Again for this to hold, $t^{-1} \lim_{t \rightarrow \infty} (\Omega \otimes t \Sigma^{22t})$ must provide a good approximation to the asymptotic variance.

Finally, the assumption in (c) implies that $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) = 0$. Using the same arguments as above this will be the case if $v_{2i1} + \dots + v_{2it}$ is stationary with probability one such that the individual-specific long-run variances of v_{2it} equal zero with probability one.

It is clear that Result 1 depends on the regression error η_{0it} being stationary when viewed as a time series. If η_{0it} contains individual-specific terms that are non-stationary over time then the cross-section variance of η_{0it} will be time-dependent. When η_{0it} contains individual-specific $I(1)$ trends when viewed as a time series, then the cross-section variance of η_{0it} normalized by t converges to a mean long-run variance as $t \rightarrow \infty$. Under assumption (b) in Result 1 this means that the asymptotic variance of $\hat{\gamma}_{2,N,t}$ converges to the ratio between two mean long-run variances as $t \rightarrow \infty$. In Section 4 we will see that this situation is similar to the one where the cointegrating parameters differ across cross-section units.

As a special case of the model defined in Section 2 we consider the following model

$$Y_{it} = \gamma'_2 X_{2it} + \mu_{0i} + u_{0t} + v_{0it} \quad (20)$$

$$X_{2it} = X_{2i0} + \mu_2 t + u_{21} + \dots + u_{2t} + v_{2i1} + \dots + v_{2it} \quad (21)$$

Here the initial value X_{2i0} and cumulated process $v_{2i1} + \dots + v_{2it}$ are the individual-specific terms in the regressor X_{2it} . Consider the case where v_{2it} is white noise with $E(v_{2it}v'_{2it}) = \Sigma_2$. From the proof of Result 1 given in Appendix A.2 it follows that $\Sigma^{22t} = (\text{Var}(X_{2i0}) + t\Sigma_2)^{-1}$ implying that for any $k > 0$ the following holds $\Sigma^{22t} > \Sigma^{22(t+k)}$, i.e. $\Sigma^{22t} - \Sigma^{22(t+k)}$ is positive definite¹. Comparing with (15) this in turn implies that the cross-section estimator $\hat{\gamma}_{2,N,t+k}$ obtained at time $t+k$ is more efficient (asymptotically as $N \rightarrow \infty$) than the cross-section estimator $\hat{\gamma}_{2,N,t}$ obtained at time t . Therefore as regards estimators of the long-run parameters in this model there is a gain in efficiency by using a cross section obtained at a later point in time when v_{2it} is white noise. However this result relies on the time-series properties of the individual-specific terms in X_{2it} and with just one cross section it is not

¹ $A_t = \text{Var}(X_{2i0}) + t\Sigma_2$ is positive definite. Then $\Sigma^{22t} - \Sigma^{22(t+k)} = A_t^{-1} - (A_t + k\Sigma_2)^{-1} = A_t^{-1} \left(A_t^{-1} + \frac{1}{k} \Sigma_2^{-1} \right)^{-1} A_t^{-1}$ is positive definite as $\left(A_t^{-1} + \frac{1}{k} \Sigma_2^{-1} \right)^{-1}$ is positive definite.

possible to get information on these properties. So in general the result only holds when t is large as stated in Result 1 above.

Once again consider the model above when v_{2it} is white noise. Using (21) it follows that for any $0 < k < t$ the regressor X_{2it} can be expressed as the sum of X_{2it-k} and an error term.

$$\begin{aligned} X_{2it} &= X_{2it-k} + w_{it} \\ w_{it} &= \mu_2 k + u_{2t-k+1} + \dots + u_{2t} + v_{2it-k+1} + \dots + v_{2it} \end{aligned}$$

As v_{2it} is white noise, dependency between X_{2it-k} and the error term w_{it} occurs only through terms that are common to all individuals. Note that this will not be the case if X_{2it} contains an individual-specific linear trend. Using the expression in (21) the cointegrating relation in (20) can be expressed as

$$Y_{it} = \gamma'_2 X_{2it-k} + \gamma'_2 w_{it} + \eta_{0it}$$

Using the same arguments as before it is clear that a cross-section regression of Y_{it}^* on X_{2it-k}^* (the variables corrected for their cross-section sample mean) will give an unbiased estimator of γ_2 . Using (20) and (21) the following expression for ΔY_{it} is obtained

$$\Delta Y_{it} = \gamma'_2 \Delta X_{2it} + \Delta \eta_{0it} = \gamma'_2 (\mu_2 + u_{2t} + v_{2it}) + \Delta \eta_{0it}$$

Again as v_{2it} is white noise, the change in the dependent variable is independent of all lags of the regressor when both are corrected for their cross-section sample mean, i.e. ΔY_{it}^* and X_{2it-k}^* are independent. This is in fact just the conditions derived in Grunfeld (1961) for a cross-section regression of the type described above to give an unbiased estimator of long-run parameters within the framework of a partial adjustment model with stationary variables. Repeating the arguments in the proof of Proposition 1 in Appendix A.1 the estimator obtained by regressing Y_{it}^* on X_{2it-k}^* is also consistent (as $N \rightarrow \infty$) with a limiting distribution which is normal. Note that the error term $\gamma'_2 w_{it} + \eta_{0it}$ is not stationary when viewed as a time series implying that the asymptotic variance is different from the one in Proposition 1.

4 A model with heterogeneous cointegrating relations

We consider an extension of the model defined in Section 2 where the cointegrating relations are allowed to differ randomly across units. The variables are now generated by the following model

$$Y_{it} = \gamma'_{1i} X_{1it} + \gamma'_{2i} X_{2it} + \eta_{0it} \quad (22)$$

$$X_{1it} = \eta_{1it} \quad (23)$$

$$X_{2it} = X_{2it-1} + \eta_{2it} \quad (24)$$

where the time-series processes η_{0it} , η_{1it} and η_{2it} are stationary for every cross-section unit $i = 1, \dots, N$ and satisfy the assumptions in Section 2. The random coefficient matrices γ_{1i} and γ_{2i} satisfy the following assumption.

Assumption 8 The $(k_1 + k_2) \times k_0$ dimensional random variables $\gamma_i = (\gamma'_{1i}, \gamma'_{2i})'$ where $i = 1, \dots, N$ are independent and identically distributed across i with finite fourth order moment.

In addition we need the terms generating the variables Y_{it} , X_{1it} and X_{2it} to satisfy a stronger moment condition and the random coefficients to be independent of all other terms as summarized in the assumptions below.

Assumption 9 X_{2i0} , μ_i , u_t , and v_{it} all have finite fourth order moments.

Assumption 10 γ_i is independent of X_{2i0} , μ_i , u_t , and v_{it} for all $i = 1, \dots, N$ and all $t = 1, 2, \dots$.

The interpretation of this model is as described in Section 2 with the difference that the cointegrating relations are allowed to differ across individuals. Using stacked notation the regression equation in (22) can be expressed as

$$Y_{it} = \gamma'_i X_{it} + \eta_{0it} = \gamma' X_{it} + w_{it} + \eta_{0it} \quad (25)$$

where $\gamma = E(\gamma_i)$ and $w_{it} = (\gamma_i - \gamma)' X_{it}$. Note that $Y_{it} - \gamma' X_{it}$ is not a cointegrating relation as w_{it} is not stationary unless $\gamma_i = E(\gamma_i)$ with probability 1, i.e. the cointegrating relations are homogeneous almost surely. Instead γ is the mean of the cointegrating parameters. Pesaran & Smith (1995) consider a special case of the model specified above where the most important difference is that in their model the common shocks are not present such that the regressors and the regression errors are assumed to be independent. The estimator obtained by a cross-section regression of the variable Y_{it} on X_{it} when both are corrected for their cross-section sample mean is defined in equation (8). It is important to note that after having subtracted the cross-section sample mean, the regressor is not independent of the regression error in (25) unless the cointegrating relations are homogenous. However, we do have conditional mean independence such that conditional on the regressor X_{it} the regression error has mean zero. In this respect, Assumption 10 about the random coefficients being independent of the regressors is crucial and this is a well-known result from the literature on random coefficient regression models, see Hildreth & Houck (1968) and Swamy (1970). The existence of fourth order moments, Assumption 9, is used when deriving the asymptotic properties of the estimator $\hat{\gamma}_{N,t}$ given in the proposition below. We note that the common shocks u_{1t} and u_{2t} are still present in w_{it} and by that in the regression error after the cross-section sample mean has been subtracted.

Proposition 2 Under Assumption 1-6 and 8-10 the following hold:

$\hat{\gamma}_{N,t}$ is a consistent estimator of $\gamma = E(\gamma_i)$, i.e.

$$\hat{\gamma}_{N,t} \xrightarrow{P} \gamma \text{ as } N \rightarrow \infty \quad (26)$$

The limiting distribution of $\hat{\gamma}_{N,t}$ is given by

$$\sqrt{N} (\hat{\gamma}_{N,t} - \gamma) \xrightarrow{w} N(0, (I_{k_0} \otimes \Sigma_t^{-1}) \Theta_t (I_{k_0} \otimes \Sigma_t^{-1})) \text{ as } N \rightarrow \infty \quad (27)$$

The asymptotic variance can be estimated consistently by using the following results

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} \xrightarrow{P} \Sigma_t \text{ as } N \rightarrow \infty \quad (28)$$

$$\frac{1}{N} \sum_{i=1}^N \text{vec} \left(X_{it}^* (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \right) \left(\text{vec} \left(X_{it}^* (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \right) \right)' \xrightarrow{P} \Theta_t \text{ as } N \rightarrow \infty \quad (29)$$

The Proof of Proposition 2 is given in Appendix A.3. The proposition shows that in a model with heterogeneous cointegrating relations a cross-section regression performed at any point in time provides a consistent estimator of the cointegrating parameter means γ . Note that since γ_i is independent of the terms generating the regressors, then γ equals the mean long-run regression coefficient used in Phillips & Moon (1999). This will also be case when the error term is $I(1)$ such that there is no cointegration. The result in Proposition 2 is shown without imposing the assumption that the regressors and the regression error are independent of each other. As before endogeneity is allowed through the common shocks. The limiting distribution is a mixed normal since the terms u_{1t} and u_{2t} are contained in Θ_t , see Appendix A.3 for details. The asymptotic variance of the cross-section estimator can be estimated consistently by using White's heteroskedastic consistent variance estimator, see White (1980). As in the model with homogeneous cointegration, the inference procedure itself is not affected by the $I(1)$ property of the regressor but the asymptotic variance of $\sqrt{N}(\hat{\gamma}_{N,t} - \gamma)$ depends on the point in time where the cross-section is obtained. The asymptotic variance viewed as a function of the point in time is very different from that in the model with homogeneous cointegration. In fact it is similar to what we would obtain in a model with no cointegration where the error term η_{0it} is $I(1)$. This is not surprising given that the regression error in (25) is $I(1)$ when viewed as a time series. The behavior of the asymptotic variance as a function of the point in time where the cross section is obtained is not derived formally but is illustrated in a simulation study in the next section.

5 A simulation study

This section illustrates the results from the previous sections in a simulation experiment. In particular, the difference between the situation where the cointegrating parameters are the same for all cross-section units and the situation where they differ across cross-section units is investigated.

We consider the following model

$$y_{it} = \gamma_i x_{it} + u_{0t} + v_{0it} \quad (30)$$

$$x_{it} = x_{it-1} + u_{2t} + v_{2it} \quad (31)$$

with

$$v_{0it} \sim \text{iid } N(0, 1) \quad v_{2it} \sim \text{iid } N(0, 1) \quad u_{2t} \sim \text{iid } N(0, \sigma_{2u}^2)$$

and the initial values of x_{it} are all equal to zero, i.e. $x_{i0} = 0$ for $i = 1, \dots, N$. When the cross-section sample mean is subtracted from all variables in the first equation above, the term u_{0t} will not appear in the equation and consequently it is omitted in the simulations of the model.

We consider three experiments:

Experiment 1: $\gamma_i = \gamma$ for $i = 1, \dots, N$

Experiment 2: $\gamma_i \sim \text{iid } N(\gamma, 2)$ and $\sigma_{2u}^2 = 0$

Experiment 3: $\gamma_i \sim \text{iid } N(\gamma, 2)$ and $\sigma_{2u}^2 = 1$

Experiment 1 is the model with homogeneous cointegrating parameters; Experiment 2 is the model with heterogeneous cointegrating parameters and no common shocks; and Experiment 3 is the model with heterogeneous cointegrating parameters and common shocks. The estimator $\hat{\gamma}_{N,t}$ is obtained by regression of y_{it} on x_{it} (both corrected for their cross-section sample mean). Since $(\hat{\gamma}_{N,t} - \gamma)$, which is the statistic of interest, does not depend on γ this parameter value can be chosen arbitrarily. Moreover, in Experiment 1 $(\hat{\gamma}_{N,t} - \gamma)$ does not depend on the common shocks u_{2t} appearing in the regressor whereas in Experiment 2 and 3 it does.

The experiments were carried out for sample sizes of $N = 100, 250, 500$ and points in time equal to $t = 1, 5, 10, 25, 100$. The results are based on 5000 replications of the model. Table 1 reports the results for the three experiments described above. In the table, *stat1* and *stat2* are defined as

$$\begin{aligned} \text{stat1} &= \sqrt{N} (\hat{\gamma}_{N,t} - \gamma) \\ \text{stat2} &= \sqrt{N} (\hat{\gamma}_{N,t} - \gamma) / \text{Est. Asymp. Stdv.} \end{aligned}$$

So *stat2* is obtained by normalizing *stat1* by a consistent estimate of its asymptotic standard deviation according to the results in Proposition 1 and 2. The empirical mean and variance of *stat1* and *stat2* are reported in columns 3-4 and 6-7, respectively. According to the asymptotic results both means should be zero, the variance of *stat1* should be equal to the number in column 4, and the variance of *stat2* should be equal to one. The asymptotic variance of *stat1* in the general case is given by

$$\text{Asymp. Var. stat1} = \text{Var}(\gamma_i) \left(\frac{E(\tilde{x}_{it}^4)}{E(\tilde{x}_{it}^2)^2} + \frac{u_{2t}^2}{E(\tilde{x}_{it}^2)} \right) + \frac{1}{E(\tilde{x}_{it}^2)} = \text{Var}(\gamma_i) \left(3 + \frac{\xi_t^2}{t} \right) + \frac{1}{t} \quad (32)$$

where $\xi_t = u_{21} + \dots + u_{2t}$ and $\tilde{x}_{it} = x_{it} - \xi_t$ such that $\tilde{x}_{it} \sim N(0, t)$. Here we have used that $E(\tilde{x}_{it}^2 x_{it}^2) = E(\tilde{x}_{it}^4) + \xi_t^2 E(\tilde{x}_{it}^2)$, see (62) in Appendix A.3. This shows that the asymptotic variance is increasing in i) the variance of the error term relative to the variance of the regressor, ii) the variance of the parameters, iii) the kurtosis coefficient of the regressor, and iv) the common component. Using the expression we have

	Asymp. Var. <i>stat1</i>	Limit as $t \rightarrow \infty$
Experiment 1	$1/t$	0
Experiment 2	$3 \text{Var}(\gamma_i) + 1/t$	$3 \text{Var}(\gamma_i)$
Experiment 3	$4 \text{Var}(\gamma_i) + 1/t$	$4 \text{Var}(\gamma_i)$

In Experiment 3 the asymptotic variance is a random variable since the term ξ_t^2/t appears. In our study this term is a χ^2 -distribution with 1 degree of freedom such that $E(\xi_t^2/t) = 1$ and the expression in the table above is obtained by using this.

Table 1: Simulation results

N	t	Mean stat1	Var stat1	Asymp. var.	Mean stat2	Var stat2
Experiment 1: $\gamma_i = \gamma$ for $i = 1, \dots, N$						
100	1	-0.013	1.021	1.000	-0.013	1.031
	5	-0.001	0.205	0.200	0.001	1.039
	10	-0.003	0.104	0.100	-0.010	1.046
	25	0.000	0.041	0.040	0.002	1.032
	100	0.001	0.010	0.010	0.014	1.037
250	1	-0.015	1.025	1.000	-0.017	1.028
	5	0.001	0.197	0.200	0.002	0.988
	10	0.002	0.101	0.100	0.005	1.016
	25	-0.002	0.039	0.040	-0.008	0.968
	100	0.002	0.010	0.010	0.019	1.011
500	1	-0.018	1.017	1.000	-0.019	1.018
	5	0.004	0.199	0.200	0.008	0.996
	10	0.001	0.100	0.100	0.002	1.005
	25	-0.004	0.040	0.040	-0.017	0.986
	100	-0.001	0.010	0.010	-0.006	1.014
Experiment 2: $\gamma_i \sim \text{iid } N(\gamma, 2)$ and $\sigma_{2u}^2 = 0$						
100	1	-0.036	6.636	7.000	-0.018	1.132
	5	-0.028	5.950	6.200	-0.012	1.177
	10	0.003	5.872	6.100	0.000	1.184
	25	-0.053	6.007	6.040	-0.025	1.194
	100	0.005	6.069	6.010	-0.001	1.212
250	1	0.000	6.967	7.000	-0.002	1.071
	5	-0.023	6.038	6.200	-0.009	1.056
	10	-0.003	5.895	6.100	0.000	1.060
	25	-0.049	5.991	6.040	-0.019	1.076
	100	-0.009	5.844	6.010	-0.009	1.051
500	1	-0.008	6.995	7.000	-0.006	1.038
	5	-0.010	6.340	6.200	-0.005	1.060
	10	0.021	6.104	6.100	0.009	1.045
	25	-0.065	6.132	6.040	-0.026	1.058
	100	0.006	6.065	6.010	0.002	1.051
Experiment 3: $\gamma_i \sim \text{iid } N(\gamma, 2)$ and $\sigma_{2u}^2 = 1$						
100	1	-0.040	8.963	9.000	-0.012	1.165
	5	-0.052	8.211	8.200	-0.015	1.207
	10	-0.014	8.184	8.100	-0.002	1.165
	25	-0.015	8.102	8.040	-0.010	1.194
	100	-0.056	7.950	8.010	-0.021	1.214
250	1	-0.009	8.925	9.000	0.004	1.057
	5	-0.018	8.146	8.200	-0.007	1.076
	10	0.035	8.274	8.100	0.013	1.096
	25	-0.004	8.236	8.040	-0.003	1.112
	100	-0.021	7.927	8.010	-0.007	1.081
500	1	-0.021	8.868	9.000	-0.005	1.027
	5	-0.009	8.138	8.200	-0.002	1.038
	10	0.045	8.444	8.100	0.015	1.072
	25	0.009	8.252	8.040	0.003	1.071
	100	0.028	8.029	8.010	0.008	1.045

The results in Table 1 show that the asymptotic distributions provide good approximations to the actual distributions of the statistics in terms of mean and variance. In Experiment 1 (homogeneous cointegration parameters), the asymptotic variance of $\sqrt{N}(\hat{\gamma}_{N,t} - \gamma)$ equals $1/t$ and hence it is decreasing in t such that there is a gain in asymptotic efficiency by using cross-section data obtained at a later point in time, see Section 3 for a detailed discussion. In Experiment 2 and 3 (heterogeneous cointegration parameters), the asymptotic variance of $\sqrt{N}(\hat{\gamma}_{N,t} - \gamma)$ decreases towards a constant as t increases. So unlike in the homogeneous case there is a lower limit of the asymptotic variance when viewed as a function of t .

6 Conclusion

This paper has specified a dynamic model in which a cross-section regression will reveal the cointegrating parameters. More specifically, we have specified a regression model describing cointegrating relations between variables at the individual level and shown that ordinary regression methods can be used in order to make asymptotic inference on the cointegrating parameters from a pure cross-section data obtained at any point in time. An important feature of the model is that the assumption about the regressors and the regression error being independent is not imposed. The model allows for some degree of correlation between the regressors and the regression error namely through aggregate shocks that are common to all cross-section units. This specification in turn provides a natural link to what is usually found in the cointegrating analysis of time-series variables at the aggregate level. The introduction of common shocks leads to the most simple type of dependency between the cross-section units and for instance individual-specific reactions to the aggregate shocks are ruled out with this formulation.

One serious drawback of having just one observation of the cross-section units at some point in time is that it is not possible to test if the time-series behavior of the variables is correctly specified. For instance it is not possible to determine whether the variables are in fact described by unit root processes. Obviously, observations of the cross-section units over time are needed in order to learn about the dynamic properties of the variables. The more observations over time the better. As mentioned in the introduction there is already papers concerning estimation of cointegrating parameters when both the cross-section dimension and the time series dimension are large. Nevertheless, it might be useful to have an idea about the framework in which it is possible to make inference on the cointegration parameters from pure cross-section data.

Acknowledgments

The author thanks M. Browning, S. Johansen and H.C. Kongsted and two anonymous referees for useful discussions and comments.

A Appendices

A.1 Proof of Proposition 1

This appendix contains the proof of Proposition 1 in the main text. The results in the proposition are all based on the Lindeberg-Levy version of the Central Limit Theorem (CLT) and the Strong Law of Large Numbers (SLLN). The lemma given below appears to be useful in following.

Lemma 1 *Consider the following regression model*

$$Y_i = \beta' X_i + \varepsilon_i \quad \text{for } i = 1, \dots, N \quad (33)$$

where the following assumptions are imposed

$$\begin{aligned} X_i \text{ and } \varepsilon_i \text{ are iid with finite second moment} \\ E(X_i) = 0 \text{ and } E(\varepsilon_i) = 0 \\ X_i \text{ and } \varepsilon_i \text{ are independent} \end{aligned} \quad (34)$$

Define the variables corrected for sample mean in the following way

$$\begin{aligned} Y_i^* &= Y_i - \frac{1}{N} \sum_{i=1}^N Y_i \\ X_i^* &= X_i - \frac{1}{N} \sum_{i=1}^N X_i \\ \varepsilon_i^* &= \varepsilon_i - \frac{1}{N} \sum_{i=1}^N \varepsilon_i \end{aligned} \quad (35)$$

The ordinary least square estimator $\hat{\beta}_N$ obtained by regressing Y_i^ on X_i^* can be expressed as*

$$\hat{\beta}_N = \beta + \left(\sum_{i=1}^N X_i^* X_i^{*'} \right)^{-1} \left(\sum_{i=1}^N X_i^* \varepsilon_i^{*'} \right) \quad (36)$$

The limiting distribution of $\sqrt{N}(\hat{\beta}_N - \beta)$ is given by the following expression

$$\sqrt{N}(\hat{\beta}_N - \beta) \xrightarrow{w} N(0, \Omega \otimes \Sigma^{-1}) \quad \text{as } N \rightarrow \infty \quad (37)$$

where $\Omega = \text{Var}(\varepsilon_i)$ and $\Sigma = \text{Var}(X_i)$. In particular $\hat{\beta}_N - \beta \xrightarrow{P} 0$ as $N \rightarrow \infty$.

Proof of Lemma 1:

First of all we show that $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^* \varepsilon_i^{*'}$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i \varepsilon_i'$ are asymptotically equivalent.

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^* \varepsilon_i^{*'} - \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i \varepsilon_i' \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\left(X_i - \frac{1}{N} \sum_{i=1}^N X_i \right) \left(\varepsilon_i - \frac{1}{N} \sum_{i=1}^N \varepsilon_i \right)' - X_i \varepsilon_i' \right) \\ &= - \left(\frac{1}{N} \sum_{i=1}^N X_i \right) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \varepsilon_i \right)' \xrightarrow{P} 0 \text{ as } N \rightarrow \infty \end{aligned}$$

since $\frac{1}{N} \sum_{i=1}^N X_i \xrightarrow{P} E(X_i) = 0$ as $N \rightarrow \infty$ by SLLN and $\frac{1}{\sqrt{N}} \sum_{i=1}^N \varepsilon_i$ converges in distribution by the Lindeberg-Levy CLT. This implies that $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^* \varepsilon_i^{*'} and $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i \varepsilon_i'$ have the same limiting distribution as given by the following expression$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^* \varepsilon_i^{*'} \xrightarrow{w} N(0, \Omega \otimes \Sigma) \text{ as } N \rightarrow \infty \quad (38)$$

This follows by the Lindeberg-Levy CLT as $X_i \varepsilon_i'$ is iid across i with $E(X_i \varepsilon_i') = 0$ and $\text{Var}(X_i \varepsilon_i') = \text{Var}(\varepsilon_i) \otimes \text{Var}(X_i) = \Omega \otimes \Sigma$ as X_i and ε_i are independent with finite second moment. Next we show that $\frac{1}{N} \sum_{i=1}^N X_i^* X_i^{*'} and $\frac{1}{N} \sum_{i=1}^N X_i X_i'$ are asymptotically equivalent.$

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N X_i^* X_i^{*'} - \frac{1}{N} \sum_{i=1}^N X_i X_i' \\ &= \frac{1}{N} \sum_{i=1}^N \left(\left(X_i - \frac{1}{N} \sum_{i=1}^N X_i \right) \left(X_i - \frac{1}{N} \sum_{i=1}^N X_i \right)' - X_i X_i' \right) \\ &= - \left(\frac{1}{N} \sum_{i=1}^N X_i \right) \left(\frac{1}{N} \sum_{i=1}^N X_i \right)' \xrightarrow{P} 0 \text{ as } N \rightarrow \infty \end{aligned}$$

since $\frac{1}{N} \sum_{i=1}^N X_i \xrightarrow{P} E(X_i) = 0$ as $N \rightarrow \infty$ by SLLN. This implies that $\frac{1}{N} \sum_{i=1}^N X_i^* X_i^{*'} and $\frac{1}{N} \sum_{i=1}^N X_i X_i'$ have the same probability limit. Using SLLN on $\frac{1}{N} \sum_{i=1}^N X_i X_i'$ the above implies the following$

$$\frac{1}{N} \sum_{i=1}^N X_i^* X_i^{*'} \xrightarrow{P} E(X_i X_i') = \text{Var}(X_i) = \Sigma \text{ as } N \rightarrow \infty \quad (39)$$

Combining (38) and (39) the following can be obtained

$$\sqrt{N} (\hat{\beta}_N - \beta) = \left(\frac{1}{N} \sum_{i=1}^N X_i^* X_i^{*'} \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^* \varepsilon_i^{*'} \right) \xrightarrow{w} N(0, \Omega \otimes \Sigma^{-1}) \text{ as } N \rightarrow \infty$$

Especially $\hat{\beta}_N - \beta \xrightarrow{P} 0$ as $N \rightarrow \infty$. \square

Proof of Proposition 1:

Summing over t in equation (3) gives the following expression for X_{2it}

$$X_{2it} = X_{2i0} + \sum_{s=1}^t \eta_{2is}$$

Inserting the expressions for η_{0it} , η_{1it} and η_{2it} given in (4)-(6) we obtain the following

$$\begin{aligned} \eta_{0it} &= \mu_{0i} + u_{0t} + v_{0it} \\ X_{1it} &= \eta_{1it} = \mu_{1i} + u_{1t} + v_{1it} \end{aligned} \quad (40)$$

$$X_{2it} = X_{2i0} + \sum_{s=1}^t (\mu_{2i} + u_{2s} + v_{2is}) = X_{2i0} + t\mu_{2i} + \sum_{s=1}^t u_{2s} + \sum_{s=1}^t v_{2is} \quad (41)$$

Subtracting the cross-section sample means in the equations above gives

$$\begin{aligned}
\eta_{0it}^* &= \mu_{0i} + v_{0it} - \frac{1}{N} \sum_{i=1}^N \mu_{0i} - \frac{1}{N} \sum_{i=1}^N v_{0it} \\
X_{1it}^* &= \mu_{1i} + v_{1it} - \frac{1}{N} \sum_{i=1}^N \mu_{1i} - \frac{1}{N} \sum_{i=1}^N v_{1it} \\
X_{2it}^* &= X_{2i0} + t\mu_{2i} + \sum_{s=1}^t v_{2is} - \frac{1}{N} \sum_{i=1}^N X_{2i0} - t \frac{1}{N} \sum_{i=1}^N \mu_{2i} - \sum_{s=1}^t \left(\frac{1}{N} \sum_{i=1}^N v_{2is} \right)
\end{aligned}$$

We define the following variables

$$\begin{aligned}
\tilde{\eta}_{0it} &= \mu_{0i} - E(\mu_{0i}) + v_{0it} \\
\tilde{X}_{1it} &= \mu_{1i} - E(\mu_{1i}) + v_{1it} \\
\tilde{X}_{2it} &= X_{2i0} - E(X_{2i0}) + t\mu_{2i} - tE(\mu_{2i}) + \sum_{s=1}^t v_{2is}
\end{aligned}$$

For every $t = 1, 2, \dots$ these variables all define sequences that are iid across i with mean zero and finite second moment. In addition \tilde{X}_{1it} and \tilde{X}_{2it} are independent of $\tilde{\eta}_{0it}$. This follows by Assumption 1-3 and 5. We define the $(k_1 + k_2)$ -dimensional stacked variable $\tilde{X}_{it} = (\tilde{X}_{1it}', \tilde{X}_{2it}')'$. For later use we need the following

$$\begin{aligned}
\Omega &\equiv \text{Var}(\tilde{\eta}_{0it}) = \text{Var}(\mu_{0i}) + \text{Var}(v_{0it}) \\
\Sigma_t &\equiv \text{Var}(\tilde{X}_{it}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12,t} \\ \Sigma'_{12,t} & \Sigma_{22,t} \end{bmatrix}
\end{aligned} \tag{42}$$

where the Σ_{ij} 's are

$$\begin{aligned}
\Sigma_{11} &\equiv \text{Var}(\tilde{X}_{1it}) = \text{Var}(\mu_{1i}) + \text{Var}(v_{1it}) \\
\Sigma_{12,t} &\equiv \text{Cov}(\tilde{X}_{1it}, \tilde{X}_{2it}) = t \text{Cov}(\mu_{1i}, \mu_{2i}) + \sum_{s=1}^t E(v_{1it} v'_{2is}) \\
\Sigma_{22,t} &\equiv \text{Var}(\tilde{X}_{2it}) = \text{Var}(X_{2i0}) + t^2 \text{Var}(\mu_{2i}) + \sum_{s=1}^t \sum_{j=1}^t E(v_{2is} v'_{2ij})
\end{aligned} \tag{43}$$

This follows by Assumptions 1-3 and 5-6. Finally note the following relations

$$\begin{aligned}
\eta_{0it}^* &= \tilde{\eta}_{0it} - \frac{1}{N} \sum_{i=1}^N \tilde{\eta}_{0it} \\
X_{1it}^* &= \tilde{X}_{1it} - \frac{1}{N} \sum_{i=1}^N \tilde{X}_{1it} \\
X_{2it}^* &= \tilde{X}_{2it} - \frac{1}{N} \sum_{i=1}^N \tilde{X}_{2it}
\end{aligned} \tag{44}$$

The estimator $\hat{\gamma}_{N,t}$ defined in (8) can be written as

$$\hat{\gamma}_{N,t} = \gamma + \left(\sum_{i=1}^N X_{it}^* X_{it}^{*'} \right)^{-1} \left(\sum_{i=1}^N X_{it}^* \eta_{0it}^{*'} \right) \tag{45}$$

According to the relations in (44) and by using Lemma 1 the estimator has a limiting distribution given by the following

$$\sqrt{N} (\hat{\gamma}_{N,t} - \gamma) \xrightarrow{w} N(0, \Omega \otimes \Sigma_t^{-1}) \text{ as } N \rightarrow \infty \tag{46}$$

In addition $\hat{\gamma}_{N,t} - \gamma \xrightarrow{P} 0$ as $N \rightarrow \infty$ meaning that $\hat{\gamma}_{N,t}$ is a consistent estimator of γ .

To show (11) we use the result in (39) from the proof of Lemma 1.

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} \xrightarrow{P} E(\tilde{X}_{it} \tilde{X}_{it}') = \text{Var}(\tilde{X}_{it}) = \Sigma_t \text{ as } N \rightarrow \infty \quad (47)$$

Using the same arguments we also obtain

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* \eta_{0it}^{*'} \xrightarrow{P} E(\tilde{X}_{it} \tilde{\eta}_{0it}') = 0 \text{ as } N \rightarrow \infty \quad (48)$$

$$\frac{1}{N} \sum_{i=1}^N \eta_{0it}^* \eta_{0it}^{*'} \xrightarrow{P} E(\tilde{\eta}_{0it} \tilde{\eta}_{0it}') = \text{Var}(\tilde{\eta}_{0it}) = \Omega \text{ as } N \rightarrow \infty \quad (49)$$

Finally, we show (12). Using that $\sum_{i=1}^N X_{it}^* (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' = 0$ and the relation $Y_{it}^* = \gamma' X_{it}^* + \eta_{0it}^*$ we obtain the following

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*) (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \\ &= \frac{1}{N} \sum_{i=1}^N Y_{it}^* (Y_{it}^{*'} - X_{it}^{*'} \hat{\gamma}_{N,t}) \\ &= \frac{1}{N} \sum_{i=1}^N (\gamma' X_{it}^* + \eta_{0it}^*) (X_{it}^{*'} \gamma + \eta_{0it}^{*'} - X_{it}^{*'} \hat{\gamma}_{N,t}) \\ &= \gamma' \frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t}) + \gamma' \frac{1}{N} \sum_{i=1}^N X_{it}^* \eta_{0it}^{*'} + \frac{1}{N} \sum_{i=1}^N \eta_{0it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t}) + \frac{1}{N} \sum_{i=1}^N \eta_{0it}^* \eta_{0it}^{*'} \end{aligned}$$

From (46) and (48) we have that as $N \rightarrow \infty$ the sequences $(\hat{\gamma}_{N,t} - \gamma)$ and $\frac{1}{N} \sum_{i=1}^N X_{it}^* \eta_{0it}^{*'}$ both converge in probability to zero. According to (47) the sequence $\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'}$ has a well-defined probability limit. This gives

$$\frac{1}{N} \sum_{i=1}^N (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*) (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' - \frac{1}{N} \sum_{i=1}^N \eta_{0it}^* \eta_{0it}^{*'} \xrightarrow{P} 0 \text{ as } N \rightarrow \infty$$

Combing this with (49) we obtain

$$\frac{1}{N} \sum_{i=1}^N (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*) (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \xrightarrow{P} \text{Var}(\tilde{\eta}_{0it}) = \Omega \text{ as } N \rightarrow \infty \quad (50)$$

which is the result in (12).

Altogether we have obtained the results stated in the proposition. \square

A.2 Proof of Result 1

This appendix contains the proof of Result 1 in the main text. The result is based on the properties of a weakly stationary process given in the lemma below.

Let C be a $k \times k$ matrix. In the following the norm of C is defined as

$$\|C\| = \max_{i,j} |C_{ij}| \quad (51)$$

Lemma 2 *Let the k -dimensional variable v_{it} be iid across i . In addition for every $i = 1, \dots, N$ the time-series process v_{it} is weakly stationary and $\Gamma_v(s)$ the corresponding mean autocovariance function. Assume that the mean autocovariances are absolutely summable i.e. $\sum_{s=0}^{\infty} \|\Gamma_v(s)\| < \infty$. In this case the following holds*

$$\frac{1}{t} \sum_{s=1}^t \sum_{j=1}^t E(v_{is} v'_{ij}) \rightarrow \sum_{s=-\infty}^{\infty} \Gamma_v(s) \text{ as } t \rightarrow \infty \quad (52)$$

For any $a > 0$ it holds that

$$\frac{1}{t^a} \sum_{s=1}^t E(v_{it} v'_{is}) \rightarrow 0 \text{ as } t \rightarrow \infty \quad (53)$$

Proof of Lemma 2:

Using that $\Gamma_v(s)$ is the mean autocovariance function of v_{it} we obtain the following

$$\sum_{s=1}^t \sum_{j=1}^t E(v_{is} v'_{ij}) = \sum_{s=1}^t \sum_{j=1}^t \Gamma_v(s-j) = t\Gamma_v(0) + \sum_{s=1}^{t-1} (t-s) (\Gamma_v(s) + \Gamma_v(s)')$$

such that

$$\frac{1}{t} \sum_{s=1}^t \sum_{j=1}^t E(v_{is} v'_{ij}) = \Gamma_v(0) + \sum_{s=1}^{t-1} (\Gamma_v(s) + \Gamma_v(s)') - \sum_{s=1}^{t-1} \frac{s}{t} (\Gamma_v(s) + \Gamma_v(s)')$$

The assumption about the autocovariances being absolutely summable implies that for any $\varepsilon > 0$ we can find j such that

$$\|\Gamma_v(j+1)\| + \|\Gamma_v(j+2)\| + \dots < \varepsilon/4$$

For this given j we can find T such that for all $t \geq T$ the following holds

$$\frac{1}{t} \|\Gamma_v(1)\| + \frac{2}{t} \|\Gamma_v(2)\| + \dots + \frac{j}{t} \|\Gamma_v(j)\| < \varepsilon/4$$

Using these results it follows that for all $t \geq T$ we have the following

$$\begin{aligned} \left\| \sum_{s=1}^{t-1} \frac{s}{t} (\Gamma_v(s) + \Gamma_v(s)') \right\| &\leq \sum_{s=1}^{t-1} \frac{s}{t} \|(\Gamma_v(s) + \Gamma_v(s)')\| \leq 2 \sum_{s=1}^{t-1} \frac{s}{t} \|\Gamma_v(s)\| \\ &= 2 \sum_{s=1}^j \frac{s}{t} \|\Gamma_v(s)\| + 2 \sum_{s=j+1}^{t-1} \frac{s}{t} \|\Gamma_v(s)\| \\ &\leq 2 \sum_{s=1}^j \frac{s}{t} \|\Gamma_v(s)\| + 2 \sum_{s=j+1}^{\infty} \|\Gamma_v(s)\| < \varepsilon \end{aligned}$$

Altogether this gives the result in (52). That is

$$\frac{1}{t} \sum_{s=1}^t \sum_{j=1}^t E(v_{is} v'_{ij}) \rightarrow \Gamma_v(0) + \sum_{s=1}^{\infty} (\Gamma_v(s) + \Gamma_v(s)') = \sum_{s=-\infty}^{\infty} \Gamma_v(s) \text{ as } t \rightarrow \infty$$

where the limit on the right hand side is well-defined as $\sum_{s=1}^{\infty} \Gamma_v(s)$ is well-defined by the assumption about the process v_{it} having absolutely summable mean autocovariances.

The result in (53) is simply a consequence of the assumption that $\sum_{s=0}^{\infty} \Gamma_v(s)$ is well-defined as the limit of $\sum_{s=0}^t \Gamma_v(s)$ as $t \rightarrow \infty$, i.e.

$$\sum_{s=1}^t E(v_{it}v'_{is}) = \sum_{s=0}^{t-1} \Gamma_v(s) \rightarrow \sum_{s=0}^{\infty} \Gamma_v(s) \text{ as } t \rightarrow \infty \quad (54)$$

□

Proof of Result 1:

Let Σ^{22t} be the lower $k_2 \times k_2$ diagonal block matrix of Σ_t^{-1} . Comparing with (42) in Appendix A.1 we obtain the following

$$\Sigma^{22t} = (\Sigma_{22,t} - \Sigma'_{12,t} \Sigma_{11}^{-1} \Sigma_{12,t})^{-1}$$

According to (43) also in Appendix A.1 we have the following expressions

$$\begin{aligned} \Sigma_{22,t} &= \text{Var}(X_{2i0}) + t^2 \text{Var}(\mu_{2i}) + \sum_{s=1}^t \sum_{j=1}^t E(v_{2is}v'_{2ij}) \\ \Sigma_{12,t} &= t \text{Cov}(\mu_{1i}, \mu_{2i}) + \sum_{s=1}^t E(v_{1it}v'_{2is}) \end{aligned} \quad (55)$$

For $a \in \mathbb{R}$ the diagonal matrix F_t is defined as

$$F_t = \begin{bmatrix} I_{k_1} & 0 \\ 0 & t^a I_{k_2} \end{bmatrix}$$

The condition in Assumption 7 concerns the limit as $t \rightarrow \infty$ of the matrix $F_t \Sigma_t F_t$ which can be decomposed in the same way as Σ_t as follows

$$F_t \Sigma_t F_t = \begin{bmatrix} \Sigma_{11} & t^a \Sigma_{12,t} \\ t^a \Sigma'_{12,t} & t^{2a} \Sigma_{22,t} \end{bmatrix} \quad (56)$$

(a) Comparing with (55) we obtain the following expressions

$$\begin{aligned} \frac{1}{t^2} \Sigma_{22,t} &= \text{Var}(\mu_{2i}) + \frac{1}{t^2} \sum_{s=1}^t \sum_{j=1}^t E(v_{2is}v'_{2ij}) + \frac{1}{t^2} \text{Var}(X_{2i0}) \\ \frac{1}{t} \Sigma_{12,t} &= \text{Cov}(\mu_{1i}, \mu_{2i}) + \frac{1}{t} \sum_{s=1}^t E(v_{1it}v'_{2is}) \end{aligned}$$

Using Lemma 2 above we obtain the following

$$\begin{aligned} \frac{1}{t^2} \Sigma_{22,t} &\rightarrow \text{Var}(\mu_{2i}) \text{ as } t \rightarrow \infty \\ \frac{1}{t} \Sigma_{12,t} &\rightarrow \text{Cov}(\mu_{1i}, \mu_{2i}) \text{ as } t \rightarrow \infty \end{aligned}$$

Setting $a = -1$ in (56) this implies that as $t \rightarrow \infty$

$$F_t \Sigma_t F_t = \begin{bmatrix} \Sigma_{11} & \frac{1}{t} \Sigma_{12,t} \\ \frac{1}{t} \Sigma'_{12,t} & \frac{1}{t^2} \Sigma_{22,t} \end{bmatrix} \rightarrow \begin{bmatrix} \Sigma_{11} & \text{Cov}(\mu_{1i}, \mu_{2i}) \\ \text{Cov}(\mu_{2i}, \mu_{1i}) & \text{Var}(\mu_{2i}) \end{bmatrix}$$

As Assumption 7 is satisfied with $a = -1$ the limit on the right hand side in the expression above is positive definite which in turn implies that $\text{Var}(\mu_{2i}) - \text{Cov}(\mu_{2i}, \mu_{1i}) \Sigma_{11}^{-1} \text{Cov}(\mu_{1i}, \mu_{2i})$ is positive definite.

This means that as $t \rightarrow \infty$

$$t^2 \Sigma^{22t} = \left(\frac{1}{t^2} \Sigma_{22,t} - \frac{1}{t} \Sigma'_{12,t} \Sigma_{11}^{-1} \frac{1}{t} \Sigma_{12,t} \right)^{-1} \rightarrow (\text{Var}(\mu_{2i}) - \text{Cov}(\mu_{2i}, \mu_{1i}) \Sigma_{11}^{-1} \text{Cov}(\mu_{1i}, \mu_{2i}))^{-1}$$

Therefore $\lim_{t \rightarrow \infty} (\Omega \otimes t^2 \Sigma^{22t})$ is well-defined which gives (17) in Result 1. Note that (16) in Assumption 7 also implies that $\text{Var}(\mu_{2i})$ is positive definite.

(b) Setting $a = -1/2$ in (56) leads us to consider the following

$$\begin{aligned} \frac{1}{t} \Sigma_{22,t} &= t \text{Var}(\mu_{2i}) + \frac{1}{t} \sum_{s=1}^t \sum_{j=1}^t E(v_{2is} v'_{2ij}) + \frac{1}{t} \text{Var}(X_{2i0}) \\ \frac{1}{t^{1/2}} \Sigma_{12,t} &= t^{1/2} \text{Cov}(\mu_{1i}, \mu_{2i}) + \frac{1}{t^{1/2}} \sum_{s=1}^t E(v_{1it} v'_{2is}) \end{aligned}$$

According to Lemma 2 the limits as $t \rightarrow \infty$ of these expressions are well-defined if and only if $\text{Var}(\mu_{2i}) = 0$. In the case where $\text{Var}(\mu_{2i}) = 0$ the results in Lemma 2 give the following as $t \rightarrow \infty$

$$\begin{aligned} \frac{1}{t} \Sigma_{22,t} &\rightarrow \sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) \\ \frac{1}{t^{1/2}} \Sigma_{12,t} &\rightarrow 0 \end{aligned}$$

where $\Gamma_{2v}(s)$ is the autocovariance function corresponding to v_{2it} . Thus we have

$$F_t \Sigma_t F_t = \begin{bmatrix} \Sigma_{11} & \frac{1}{t^{1/2}} \Sigma_{12,t} \\ \frac{1}{t^{1/2}} \Sigma'_{12,t} & \frac{1}{t} \Sigma_{22,t} \end{bmatrix} \rightarrow \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) \end{bmatrix}$$

Again as Assumption 7 is satisfied with $a = -1/2$ the limit on the right hand side in the expression above is positive definite which in particular means that $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s)$ is positive definite. When the autocovariances are bounded such that $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) = E(\sum_{s=-\infty}^{\infty} \Gamma_{2v}(i, s))$ then the result means that positive definite if the individual-specific long-run variances $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(i, s)$ are positive definite with probability 1, if and only if the mean long-run variance is positive definite. This gives that as $t \rightarrow \infty$

$$t \Sigma^{22t} = \left(\frac{1}{t} \Sigma_{22,t} - \frac{1}{t^{1/2}} \Sigma'_{12,t} \Sigma_{11}^{-1} \frac{1}{t^{1/2}} \Sigma_{12,t} \right)^{-1} \rightarrow \left(\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) \right)^{-1}$$

which gives (18) in Result 1. Note that the limit of $\frac{1}{t} \Sigma_{22,t}$ is the mean long-run variance of the process v_{2it} .

(c) When $a = 0$ we have $F_t = I_{k_1+k_2}$. In this case the assumption that (16) is satisfied implies that $\lim_{t \rightarrow \infty} (\Sigma_t)$ is well-defined which in turn implies that $\lim_{t \rightarrow \infty} (\Sigma^{22t})$ is well-defined. This gives (19) in Result 1. Consider the expressions in (55). One necessary condition for these expressions to have well-defined limits as $t \rightarrow \infty$ is that $\text{Var}(\mu_{2i}) = 0$. Another necessary condition is that $\frac{1}{t} \sum_{s=1}^t \sum_{j=1}^t E(v_{2is} v'_{2ij}) \rightarrow 0$ as $t \rightarrow \infty$ that is $\sum_{s=-\infty}^{\infty} \Gamma_{2v}(s) = 0$. Otherwise the limit as $t \rightarrow \infty$ of $\sum_{s=1}^t \sum_{j=1}^t E(v_{2is} v'_{2ij})$ is not well-defined. \square

A.3 Proof of Proposition 2

This appendix contains the proof of Proposition 2 in the main text. As before the proofs are based on the Lindeberg-Levy CLT and SLLN.

Proof of Proposition 2:

From (47) and (48) in Appendix A.1 we have the following results

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} \xrightarrow{P} \Sigma_t \text{ as } N \rightarrow \infty \quad (57)$$

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* \eta_{0it}^{*'} \xrightarrow{P} 0 \text{ as } N \rightarrow \infty \quad (58)$$

The estimator $\hat{\gamma}_{N,t}$ can be expressed as follows

$$\hat{\gamma}_{N,t} = \gamma + \left(\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} \right)^{-1} \left(\left(\frac{1}{N} \sum_{i=1}^N X_{it}^* w_{it}^{*'} \right) + \left(\frac{1}{N} \sum_{i=1}^N X_{it}^* \eta_{0it}^{*'} \right) \right) \quad (59)$$

where $w_{it} = (\gamma_i - \gamma)' X_{it}$ and $w_{it}^* = w_{it} - \frac{1}{N} \sum_{i=1}^N w_{it}$. Thus $\hat{\gamma}_{N,t}$ is a consistent estimator of γ if $\frac{1}{N} \sum_{i=1}^N X_{it}^* w_{it}^{*'} \xrightarrow{P} 0$ as $N \rightarrow \infty$. To show this we use the results below. As γ_i is independent of X_{2i0} , μ_i and v_{it} by Assumption 10, SLLN gives the following as $N \rightarrow \infty$

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (\gamma_{2i} - \gamma_2)' X_{2i0} &\xrightarrow{P} E((\gamma_i - \gamma)' X_{2i0}) = E(\gamma_i - \gamma)' E(X_{2i0}) = 0 \\ \frac{1}{N} \sum_{i=1}^N (\gamma_{2i} - \gamma_2)' \eta_{2it} &= \frac{1}{N} \sum_{i=1}^N (\gamma_{2i} - \gamma_2)' (\mu_{2i} + u_{2t} + v_{2it}) \xrightarrow{P} E(\gamma_{2i} - \gamma_2)' (E(\mu_{2i}) + u_{2t} + E(v_{2it})) = 0 \\ \frac{1}{N} \sum_{i=1}^N (\gamma_{1i} - \gamma_1)' \eta_{1it} &= \frac{1}{N} \sum_{i=1}^N (\gamma_{1i} - \gamma_1)' (\mu_{1i} + u_{1t} + v_{1it}) \xrightarrow{P} 0 \end{aligned}$$

Comparing with the expressions for X_{1it} and X_{2it} in (40) - (41) in Appendix A.1, the above implies that

$$\frac{1}{N} \sum_{i=1}^N w_{it} = \frac{1}{N} \sum_{i=1}^N (\gamma_i - \gamma)' X_{it} \xrightarrow{P} 0 \text{ as } N \rightarrow \infty \quad (60)$$

By using this and $\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} \xrightarrow{P} 0$ as $N \rightarrow \infty$ it follows that $\frac{1}{N} \sum_{i=1}^N X_{it}^* w_{it}^{*'} and $\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} w_{it}'$ are asymptotically equivalent, i.e.$

$$\frac{1}{N} \sum_{i=1}^N X_{it}^* w_{it}^{*'} - \frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} w_{it}' = - \left(\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} \right) \left(\frac{1}{N} \sum_{i=1}^N w_{it} \right) \xrightarrow{P} 0 \text{ as } N \rightarrow \infty$$

This implies that $\frac{1}{N} \sum_{i=1}^N X_{it}^* w_{it}^{*'} and $\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} w_{it}'$ have the same probability limit as $N \rightarrow \infty$. Therefore showing that $\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} w_{it}' \xrightarrow{P} 0$ as $N \rightarrow \infty$ implies that $\hat{\gamma}_{N,t}$ is a consistent estimator of γ . Conditional on u_{1t} and u_{2t} the terms w_{it} where $i = 1, \dots, N$ are iid across i such that the same holds for $\tilde{X}_{it} w_{it}'$. Then SLLN gives$

$$\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} w_{it}' \xrightarrow{P} 0 \text{ as } N \rightarrow \infty$$

as $E(X_{it} w_{it}') = E(X_{it} X_{it}' (\gamma_i - \gamma)) = E(X_{it} X_{it}') E(\gamma_i - \gamma) = 0$ since both X_{it} and \tilde{X}_{it} are independent of γ_i by Assumption 10.

Again conditional on u_{1t} and u_{2t} the terms $w_{it} = (\gamma_i - \gamma)' X_{it}$ where $i = 1, \dots, N$ are iid across i . First of all we show that $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_{it}^* (w_{it}^* + \eta_{0it}^*)'$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})'$ are asymptotically equivalent. Using the relations in (44) in Appendix A.1 the following is obtained

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_{i=1}^N X_{it}^* (w_{it}^* + \eta_{0it}^*)' - \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})' \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\tilde{X}_{it} - \frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} \right) \left(w_{it} - \frac{1}{N} \sum_{i=1}^N w_{it} + \tilde{\eta}_{0it} - \frac{1}{N} \sum_{i=1}^N \tilde{\eta}_{0it} \right)' - \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})' \\ &= - \left(\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} \right) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N w_{it} \right)' - \left(\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} \right) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\eta}_{0it} \right)' \xrightarrow{P} 0 \text{ as } N \rightarrow \infty \end{aligned}$$

since $\frac{1}{N} \sum_{i=1}^N \tilde{X}_{it} \xrightarrow{P} E(\tilde{X}_{it}) = 0$ as $N \rightarrow \infty$ by SLLN and both $\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\eta}_{0it}$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N w_{it}$ converge in distribution by the Lindeberg-Levy CLT as the terms $\tilde{\eta}_{0it}$ and w_{it} where $i = 1, \dots, N$ are iid across i with finite second order moments. This implies that $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_{it}^* (w_{it}^* + \eta_{0it}^*)'$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})'$ have the same limiting distribution as given by the expression below

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N X_{it}^* (w_{it}^* + \eta_{0it}^*)' \xrightarrow{w} N(0, \Theta_t) \text{ as } N \rightarrow \infty \quad (61)$$

where $\Theta_t = \text{Var}(\tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})')$. This follows by the Lindeberg-Levy CLT as $\tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})'$ is iid across i with mean

$$E(\tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})') = E(\tilde{X}_{it} X_{it}' (\gamma_i - \gamma)) + E(\tilde{X}_{it} \tilde{\eta}_{0it}') = 0$$

and variance

$$\begin{aligned} \Theta_t &= \text{Var}(\tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})') \\ &= \text{Var}(\tilde{X}_{it} w_{it}') + \text{Var}(\tilde{X}_{it} \tilde{\eta}_{0it}') \\ &= \text{Var}(\tilde{X}_{it} X_{it}' (\gamma_i - \gamma)) + \text{Var}(\tilde{X}_{it} \tilde{\eta}_{0it}') \\ &= E\left(\text{vec}(\tilde{X}_{it} X_{it}' (\gamma_i - \gamma)) \left(\text{vec}(\tilde{X}_{it} X_{it}' (\gamma_i - \gamma))\right)'\right) + \text{Var}(\tilde{\eta}_{0it}) \otimes \text{Var}(\tilde{X}_{it}) \\ &= E\left(\left(I_{k_0} \otimes \tilde{X}_{it} X_{it}'\right) \text{vec}(\gamma_i - \gamma) (\text{vec}(\gamma_i - \gamma))' \left(I_{k_0} \otimes X_{it} \tilde{X}_{it}'\right)\right) + \Omega \otimes \Sigma_t \\ &= E\left(\left(I_{k_0} \otimes \tilde{X}_{it} X_{it}'\right) \text{Var}(\gamma_i - \gamma) \left(I_{k_0} \otimes X_{it} \tilde{X}_{it}'\right)\right) + \Omega \otimes \Sigma_t \end{aligned} \quad (62)$$

The second line in the expression for the variance above results from $\tilde{X}_{it} w_{it}'$ and $\tilde{X}_{it} \tilde{\eta}_{0it}'$ being uncorrelated as $\tilde{\eta}_{0it}$ is independent of \tilde{X}_{it} and w_{it} and has mean zero. From the last line in the expression above it is clear that the variance of $\tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})'$ is well-defined as X_{it} has finite fourth moments by Assumption 9. Now combining (57) and (61) the following is obtained

$$\begin{aligned} \sqrt{N}(\hat{\gamma}_{N,t} - \gamma) &= \left(\frac{1}{N} \sum_{i=1}^N X_{it}^* X_{it}^{*'} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N X_{it}^* (w_{it}^* + \eta_{0it}^*)' \\ &\xrightarrow{w} N(0, (I_{k_0} \otimes \Sigma_t^{-1}) \Theta_t (I_{k_0} \otimes \Sigma_t^{-1})) \text{ as } N \rightarrow \infty \end{aligned}$$

which is the result (27) in Proposition 2.

Finally to show (29) we use the following

$$\frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'}) (\text{vec} (X_{it}^* X_{it}^{*'}))' = O_P(1) \quad (63)$$

$$\frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'}) (\text{vec} (X_{it}^* \eta_{0it}^{*'}))' = O_P(1) \quad (64)$$

$$\frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'}) (\text{vec} (X_{it}^* w_{it}^{*'}))' = O_P(1) \quad (65)$$

This follows as all the variables \tilde{X}_{it} , $\tilde{\eta}_{0it}$ and w_{it} are iid across i with finite fourth order moments. Using these results together with $(\hat{\gamma}_{N,t} - \gamma) \xrightarrow{P} 0$ as $N \rightarrow \infty$ give

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \text{vec} \left(X_{it}^* (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \right) \left(\text{vec} \left(X_{it}^* (Y_{it}^* - \hat{\gamma}'_{N,t} X_{it}^*)' \right) \right)' \\ & - \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \\ = & \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t}) + X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t}) + X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \\ & - \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \\ = & \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t})) (\text{vec} (X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t})))' \\ & + \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t})) (\text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \\ & + \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* X_{it}^{*'} (\gamma - \hat{\gamma}_{N,t})))' \\ = & \left((\gamma - \hat{\gamma}_{N,t})' \otimes I_{k_1+k_2} \right) \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'}) (\text{vec} (X_{it}^* X_{it}^{*'}))' ((\gamma - \hat{\gamma}_{N,t}) \otimes I_{k_1+k_2}) \\ & + \left((\gamma - \hat{\gamma}_{N,t})' \otimes I_{k_1+k_2} \right) \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* X_{it}^{*'}) (\text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \\ & + \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* X_{it}^{*'}))' ((\gamma - \hat{\gamma}_{N,t}) \otimes I_{k_1+k_2}) \xrightarrow{P} 0 \text{ as } N \rightarrow \infty \end{aligned}$$

What remains to be show is that $\frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \xrightarrow{P} \Theta_t$ as $N \rightarrow \infty$. By using similar results as above and that $\tilde{X}_{it} w_{it}'$ and $\tilde{X}_{it} \tilde{\eta}_{0it}'$ are uncorrelated, we obtain the

following

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N \text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}) (\text{vec} (X_{it}^* w_{it}^{*'} + X_{it}^* \eta_{0it}^{*'}))' \xrightarrow{P} \\
& E \left(\text{vec} (\tilde{X}_{it} w_{it}') (\text{vec} (\tilde{X}_{it} w_{it}'))' \right) + E \left(\text{vec} (\tilde{X}_{it} \tilde{\eta}_{0it}') (\text{vec} (\tilde{X}_{it} \tilde{\eta}_{0it}'))' \right) \\
& + E \left(\text{vec} (\tilde{X}_{it} w_{it}') (\text{vec} (\tilde{X}_{it} \tilde{\eta}_{0it}'))' \right) + E \left(\text{vec} (\tilde{X}_{it} \tilde{\eta}_{0it}') (\text{vec} (\tilde{X}_{it} w_{it}'))' \right) \\
& = E \left(\text{vec} (\tilde{X}_{it} w_{it}') (\text{vec} (\tilde{X}_{it} w_{it}'))' \right) + E \left(\text{vec} (\tilde{X}_{it} \tilde{\eta}_{0it}') (\text{vec} (\tilde{X}_{it} \tilde{\eta}_{0it}'))' \right) \\
& = \text{Var} (\tilde{X}_{it} (w_{it} + \tilde{\eta}_{0it})') = \Theta_t
\end{aligned}$$

Altogether we have shown the results stated in Proposition 2. \square

References

- Adda, J. and J-M. Robin, 1998, Estimation from cross-sections of integrated time-series, Working Paper n 9802, C.E.P.R.E.M.A.P., France.
- Adda, J. and J-M. Robin, 2003, Aggregation of non stationary demand systems, Contributions to Economic Analysis & Policy, Volume 2, Issue 1, Article 7.
- Arellano, M., 2003, Panel data econometrics, (Oxford University Press, Oxford).
- Baltagi, B.H., 2000, Advances in Econometrics vol.15, Nonstationary panels, panel cointegration, and dynamic panels, (Elsevier, New York).
- Baltagi, B.H. and C. Kao, 2000, Nonstationary panels, cointegration in panels and dynamic panels: A survey, in: Baltagi, B.H. (Ed), Advances in Econometrics vol. 15, 7-52.
- Baltagi, B., 2001, Econometric analysis of panel data, Second edition (Wiley, New York).
- Banerjee, A., 1999, Panel data unit roots and cointegration: An overview, Oxford Bulletin of Economics and Statistics, Special Issue, 607-630.
- Grunfeld, Y., 1961, The interpretation of cross section estimates in a dynamic model, *Econometrica* 29, 397-405.
- Hildreth, C. and J.P. Houck, 1968, Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association* 63, 584-595.
- Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).
- Mátyas, L. and P. Sevestre, 1992, The econometrics of panel data (Kluwer Academic Publishers, Boston, Massachusetts).
- Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, *Econometric Theory* 5, 95-131.
- Pesaran, M.H. and R. Smith, 1995, Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics* 68, 79-113.
- Phillips, P.C.B., 1986, Understanding spurious regressions in econometrics, *Journal of Econometrics* 33, 311-40.
- Phillips, P.C.B. and H.R. Moon, 1999, Linear regression limit theory for nonstationary panel data, *Econometrica* 67, 1057-1111.
- Swamy, P.A.V.B., 1970, Efficient inference in a random coefficient regression model, *Econometrica* 38, 311-323.
- White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-838.